CRISPR disruption and UK Biobank analysis of a highly conserved polymorphic enhancer suggests its role in anxiety and male alcohol intake.

Andrew McEwan[1], Elizabeth Hay[1], Connor Davidson[1], Yvonne Turnbull[1], Pietro Marini[1], Dana Wilson[1], Andrew M. McIntosh[3,4] Mark J. Adams[4], Chris Murgatroyd[2], Perry Barrett[1], Mirela Delibegovic[1], Toni-Kim Clarke[4] and Alasdair MacKenzie[1*].

1. School of Medicine, Medical Sciences and Nutrition, Institute of Medical Sciences, Foresterhill, University of Aberdeen, Aberdeen, Scotland, UK, AB25 2ZD.

2. Manchester Metropolitan University, Manchester UK.

3. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK

4. Division of Psychiatry, University of Edinburgh, UK

*Corresponding Author. School of Medicine, Medical Sciences and Nutrition, Institute of Medical Sciences Foresterhill, University of Aberdeen, Aberdeen, Scotland, UK, AB25 2ZD, Tel; +44(0)1224437380. e-mail; mbi167@abdn.ac.uk.

Running Title; CRISPR analysis of an enhancer controlling ethanol intake.

Key Words. UK Biobank, CRISPR genome editing, tissue-specific enhancer, galanin, ethanol intake, anxiety.

Abstract.

Excessive alcohol intake is associated with 5.9% of global deaths. However, this figure is especially acute in men such that 7.6% of deaths can be attributed to alcohol intake against only 4% in women. Previous studies identified a significant interaction between haplotypes of the GAL gene with anxiety and alcohol abuse in different male populations but were unable to define a mechanism. To address these issues the current study analysed the human UK Biobank cohort and identified a significant interaction (n=115,865; *p*=0.0008) between allelic variation (GG or CA haplotypes) in the highly conserved human GAL5.1 enhancer, alcohol intake (AUDIT questionnaire scores) and anxiety in men that was consistent with these previous studies. Critically, disruption of GAL5.1 in mice using CRISPR genome editing significantly and specifically reduced GAL expression in amygdala and hypothalamus whilst producing a corresponding reduction in ethanol intake in KO mice. Intriguingly, we also found reduced anxiety in male animals lacking GAL5.1 demonstrating sexual dimorphism mirroring that seen in humans. Using bioinformatic analysis and co-transfection studies we further identified the EGR1 transcription factor, that is co expressed with GAL in amygdala and hypothalamus, as being important in the PKC supported activity of the GG haplotype of GAL5.1 but less so in the CA haplotype. Our unique study uses a novel combination of human association analysis, CRISPR genome editing in mice, animal behavioural analysis and cell culture studies to identify a highly conserved regulatory mechanism linking anxiety and alcohol intake that might contribute to anxiety induced alcohol abuse in men.

Background.

The relationship between alcohol abuse and anxiety has been extensively explored [1] but the genomic mechanisms linking them remain poorly understood. Previous studies have shown that the galanin neuropeptide (encoded by the *GAL* gene) influences alcohol intake [2, 3] and mood [4]. Moreover, genetic analyses of different haplotypes within the GAL locus succeeded in identifying an association with excess alcohol intake that was influenced by sex and axiety[5]. Unfortunately, these studies were unable to define a mechanism to explain these interactions as the GAL coding region lacks non-synonymous polymorphisms[5]. Although there is a possibility that mis-regulation of GAL may affect alcohol intake and anxiety, little is known of the genomic mechanisms that modulate the expression of the GAL gene in the brain.

Based on the hypothesis that many regions of the genome that control gene expression are essential to species fitness are conserved through evolution [6, 7], we undertook a comparative genomic analysis of the genome surrounding the human GAL locus and succeeded in identifying a highly conserved enhancer sequence (hGAL5.1) 42 kilobases (kb) from the GAL gene transcriptional start site. We demonstrated that GAL5.1 was active in galanin expressing cells of the hypothalamus including the PVN and dorsomedial hypothalamus (DMH)[8]. We also found that GAL5.1 contained two polymorphisms in perfect linkage disequilibrium (LD; rs2513280 (C/G) and rs2513281 (A/G)) that produced two different haplotypes (GG and CA) within the human population and reported that the major GG haplotype was significantly more active in primary hypothalamic neurones than the minor CA haplotype [8]. Subsequent analysis of these haplotypes in human populations detected an association of the GG haplotype and volume of alcohol intake in women in a small US cohort (n=138) [9]. However, a second UK (n=2731) and US cohort (n=4064) based study, although initially identifying an association to increased frequency of binge drinking in teenagers, failed to reach significance after correction for multiple comparison [10]. Because of the comparatively small size of these studies we interrogated a much larger human cohort (UK biobank) to better determine the association of specific allelic variants in GAL5.1 with levels of alcohol intake and anxiety in UK men and women. In addition, we used cutting edge CRISPR genome editing to

disrupt GAL5.1 in mice and examined the effects on the expression of flanking genes in areas of the brain including the hypothalamus and the amygdala to functionally link GAL5.1 activity to the expression of five flanking genes including *Gal.* Our study also characterises the effects of GAL5.1 disruption on ethanol consumption and anxiety behaviour in mice. A combination of bioinformatics (DNAseI hypersensitivity) and co-transfection studies also identified a transcription factor-DNA interaction underpinning GAL5.1 function and its response to signal transduction cues. This study highlights a mechanistic link between ethanol consumption and anxiety centred on the GAL5.1 enhancer that may contribute to the development of alcohol abuse and anxiety in men[11, 12].

Methods.

Genetic Association Study. We first analysed the association of alcohol intake and allelic variation at the rs2513280 locus in a population of 345,140 individuals (UK Biobank; 183,921 females and 161,219 males). In the MHQ, alcohol use was measured using the AUDIT (Babor, 1991: The Alcohol Use Disorders Identification Test – Guidelines for Use in Primary Care. World Health Organ - Dep Ment Health Subst Depend 2001); a 10-item questionnaire used to measure both alcohol consumption and problematic drinking. AUDIT total scores range from 0-40 and the derivation of this measure in the UKB has been described previously in greater detail [13].

To analyse the interaction effect of rs2513280 and anxiety on alcohol use behaviour in humans we also analysed a subset of the UK Biobank[14] who had responded to a mental health questionnaire (MHQ the results of which were made available to researchers in August 2017 [15]. Anxiety was ascertained by asking participants "Over the last 2 weeks, how often have you been bothered by any of the following problems? Feeling nervous, anxious or on edge" [UKB data field 20506]. Anxiety was then dichotomized by comparing those answering 'Not at all' to individuals endorsing 'Several days', 'More than half the days' or 'Every day'. After filtering the MHQ subset of individuals on those who were White, British and unrelated with and non-missing data there remained 115,865 individuals available for analysis. Using AUDIT score as the outcome variable we performed linear regression using age fitting 20 genetics principal components as fixed effect covariates. rs2513280 was fit as a main effect (coded 0,1,2 corresponding to the number of C alleles carried) alongside anxiety and sex and then fitting a three-way interaction term of sex*anxiety*rs2513280.

Generation of gRNA molecules by a novel annealed oligo template (AOT) method. Single guide RNA (sgRNA) molecules were designed to disrupt the GAL5.1 enhancer (Fig. 1) using the optimised CRISPR design tool (http://CRISPR.mit.edu/). sgRNA template was produced by annealing two oligonucleotides. The first was a generic 80mer (5'-AAA AAA AGC ACC GAC TCG GTG CCA CTT TTT CAA GTT GAT AAC GGA CTA GCC TTA TTT TAA *CTT*

GCT ATT TCT AGC TCT AA-3') that represented the sgRNA scaffold and was called the template scaffold oligo. The second oligonucleotide was termed the sgRNA specific oligo (63mer; 5'-TTA ATA CGA CTC ACT ATA GGN NNN NNN NNN NNN NNN NNN GTT *TTA GAG CTA GAA ATA GCA AG* -3') where "N" denotes the predicted guide sequence (5'sgRNA; CTC CCT GGA GCA ATA TGA AG and 3'sgRNA; CCC GCT TTC ATG GCT CCC AA). Underlined sequence represents the T7 promoter and italic sequence represents sequence complimentary to the template scaffold oligo. These oligonucleotides were annealed and amplified using PCR to produce a 122 bp double strand sgRNA template. 100 ng of this template was used to produce sgRNA using a mMESSAGE mMACHINE T7 *in-vitro* transcription kit (Ambion) described in the manufacturer's instructions and purified using a Megaclear kit (Ambion) with modifications as previously described [16].

Production of genome edited mice. sgRNA molecules were microinjected at a concentration 10 ng/µl each into the cytoplasm of 1-cell C57/BL6 embryos as described [16] together with 10 ng/µl CAS9 mRNA (Life Technologies). Two-cell embryos were introduced into host CD1 mothers using oviduct transfer as previously described [17] and correctly targeted offspring were determined by PCR of earclip DNA using the following flanking primers (mGAL5for; AGTTAGGGCGCACACATCAA, mGAL5rev; CCGTGACTAACG GCTAATGC).

***In-situ* Hybridisation.** Radioactive *in-situ* hybridisation was carried out on 10µm brain sections derived from wild type mice using radiolabelled *Gal* or *EGR1* antisense RNA probes as previously described [18].

Quantitative reverse transcriptase-PCR. Brain tissues (Hypothalamus, amygdala, hippocampus and cortex) were dissected and snap frozen on dry ice. Total RNA was extracted using the isolate II RNA minikit (Bioline). Quantitative reverse transcriptase-PCR (QrtPCR) to determine mRNA expression levels of genes flanking mGAL5.1 (*Lrp5, Ppp6r3, Gal, Mlt5 and Cpt1a*) was undertaken using gene specific primers (See table 1 in supplementary data) on derived cDNA using mouse Qrt-PCR primers as previously described [19, 20] using a Roche Light Cycler 480 with Roche SYBR green. All QrtPCR analyses were normalised using mouse

primers specific to the *Nono* housekeeping gene that gave the most stable expression in all neuronal tissues analysed compared to βactin, HPRT or GAPDH genes.

Alcohol preference studies. All animal studies were performed in accordance with UK Home Office guidelines. All mice used were sex and age matched littermates. Oral 10% ethanol self-administration and preference were examined using a two-bottle choice protocol as described previously [21]. Mice were housed in TSE home cage systems (TSE, Bad Homburg, Germany) that record liquid intake automatically. Initially mice were group housed (~4 / cage) and habituated for 5 days to allow for adaptation to the monitored bottles. The mice were subsequently singly housed and habituated for a further 5 days prior to introduction of a second bottle containing the ethanol solution. Intake of both water and 10% ethanol solution (both containing 0.05% saccharine to normalise preference to ethanol) were monitored at 30 min intervals over the course of the trial and recorded by TSE PhenoMaster Software. Bottles were regularly swapped to reduce positional effects on consumption.

Open Field test. Anxiety like behaviour in GAL5.1KO and WT littermate mice was tested using the open field test [22]. Briefly, the open field test consisted of a square 30□cm (30□cm high) PVC open field arena positioned on a white base with overhead lighting applying 100 lx at the base. Animals were transported individually to the testing room, habituated (30 mins) and released into the corner of the arena. Ambulatory activity was recorded in the open field for 600 s using an overhead video camera and the ANY-maze tracking software. The software was used to define three zones in the open field at approximately 15 cm^2 centre zone, 20 cm^2 mid zone and 30 cm^2 peripheral zone. Distance travelled, mean speed, Peripheral zone and no. centre entries, were determined by the software.

Deletion constructs. Primers were designed that flanked the conserved putative EGR1 binding site within the GAL5.1 enhancer (ARM212-D2F; ATA GAT TTC AGA AAA GAA AGC TT, ARM213-D2R; TAA AAT GAC TGG CAT TAG AGC TC -3'). These were used in a PCR reaction with Q5 Hi-Fidelity polymerase (NEB, USA) in combination with the pLuc-GAL5.1 reporter construct previously described [8] as template to produce pLuc-GALΔEGR. The PCR

product was ligated with T4 ligase prior to transformation into competent cells (Stratagene, UK). The plasmid was sequenced to validate the removal of the putative EGR1 binding site.

Cell culture and transfection studies. SH-SY5Y cells (94030304, ECAC, UK) were cultured in Dulbecco's Modified Eagle Medium (DMEM; Gibco, UK) containing low Glucose (5.5mM), L-Glutamine (4mM), and Sodium Pyruvate (1mM). Medium was previously supplemented with 10% (v/v) heat-inactivated Foetal Bovine Serum (FBS; Gibco) and 1% (v/v) Penicillin-Streptomycin (Pen-Strep; Gibco, UK). Cells were transfected with plasmid DNA using jetPRIME as per manufacturer's instruction (Polyplus Illkirch-France). After transfection cells were treated with phorbol 12**-**myristate 13-acetate (PMA; 100nM in DMSO) or DMSO for 24 hours following which cells were lysed for dual luciferase analysis as per manufacturer's instructions (Promega UK).

Data analysis. From in-vivo pilot studies we calculated that a minimum of 6-12 animals per group would enable detection of a 25% difference between different parameters (ethanol intake, fat deposition) with 80% power using one-way ANOVA and/or general linear modelling. Statistical significance of data sets was analysed using either one-way analysis of variance (ANOVA) analysis with Bonferroni post hoc tests or using one tailed or two tailed unpaired parametric Student *t*-test as indicated using GraphPad PRISM version 5.02 (GraphPad Software, La Jolla, CA, USA).

Results.

The G-allele of rs2513280 associates with increased alcohol consumption in women. In order to test the validity of two previous association analyses[9, 10] we investigated whether there was an association between allelic variants of the GAL5.1 enhancer and increased alcohol intake in a much larger cohort comprising 345,140 individuals (UK Biobank; 183,921 females and 161,219 males). In the total sample an association between the G allele of rs2513280 (a proxy for both loci) and increased weekly alcohol intake was identified (b=0.008, s.e.=0.004, p=0.0463). When analysing males and females separately, the association between male alcohol intake and rs2513280 became non-significant (b=0.004, s.e.=0.006, p=0.478); however, in females the association was stronger and remained significant (b=0.012, s.e.=0.006, p=0.037).

Because of the known link between anxiety and alcohol abuse in men[5] we further investigated the interaction of these variables with respect to the rs2513280 using a subset of the UK Biobank that also contained information relating to anxiety[15]. There was an effect of male sex on AUDIT score with males having significantly higher AUDIT scores than females (b=0.34, S.E.=0.005, $p < 2 \times 10^{-16}$). Individuals with higher levels of self-reported anxiety also had significantly higher AUDIT scores although there was no main effect of the rs2513280 polymorphism. No significant interactions were detected between sex and rs2153280 or anxiety and rs2513280 (**Fig 1A**). However, anxious males reported significantly higher AUDIT scores (*p*=0.0008) if they carried the major allele at the rs2153280 locus (**Fig 1A and B**).

The GAL5.1 enhancer supports reporter gene expression in specific regions of the brain including the hypothalamus and amygdala. Cloning of the human GAL5.1 enhancer next to a LacZ reporter (hGAL5.1-LacZ) and production of transgenic reporter mice demonstrated expression of the LacZ reporter in cells of the periventricular (PVN), dorsomedial (DMH) and arcuate nuclei (ARC; **Fig 2B**) of the hypothalamus as well as the medial nucleus of the amygdala (MeA; **Fig 2C**) that corresponded to the expression patterns of the endogenous wild type mouse *Gal* gene both at a tissue-specific (**Fig2A and C**) and cellular level [8].

Rapid and accurate disruption of GAL5.1 in mice was achieved using CRISPR/CAS9. To determine a functional role for GAL5.1 in-vivo we chose two target sequences (5'sgRNA and 3'sgRNA) that flanked the most highly conserved region of mGAL5.1 (chr19:3440931-3441185; **Fig. 2E**) to produce a disruptive 230 bp deletion. We devised a novel and highly efficient method that we termed the annealed oligo template (AOT) method that generates a single guide RNA (sgRNA) template from DNA oligonucleotide precursors. We microinjected AOT derived sgRNA and CAS9mRNA into the cytoplasm of 90 1-cell mouse embryos. 90% of these embryos survived and were oviduct transferred into host female CD1 mice to generate a homozygous female and two heterozygous male mice that contained identical deletions within GAL5.1 following analysis by PCR and electrophoresis (mGAL5.1KO; **Fig. 2F and G**) without any evidence of off-target effects [23].

Disruption of GAL5.1 reduces expression of GAL throughout the brain but does not affect the expression of flanking genes. We compared the expression of the GAL gene in mRNA derived from various regions of the brain in wild type and mGAL5.1KO mice. Expression of GAL mRNA was strongest in the hypothalamus and the amygdala with some evidence of GAL mRNA expression in hippocampus and cortex (**Fig 3A and B**). We observed that expression of *Gal* became almost undetectable in all these tissues in GAL5.1KO animals (**Fig 3A and B)** suggesting that GAL5.1 is essential for global *Gal* expression.

We also analysed the influence of mGAL5.1 disruption on the expression of *Gal* and 4 other flanking genes flanking the Gal locus namely *Lrp5, Ppp6r3* (5' of GAL5.1) *Mtl5* and *Cpt1a* (3'of GAL5.1) that are all maintained in the same synteny block [24] in both humans and mice spanning 536kb in humans and 379kb in mice (**Fig 3C**). Although we found strong down regulation of the Gal gene we saw no significant change of expression in any of the flanking genes suggesting a lack of requirement for GAL5.1 for expression of these genes (**Fig 3D**)[25].

mGAL5.1KO animals exhibit a decreased preference for ethanol.

Because the GAL gene has been shown to control intake of alcohol[2, 3] and that a polymorphism within the GAL5.1 enhancer was associated with increased alcohol intake in women we tested the hypothesis that disruption of the mGAL5.1 enhancer would reduce

preference for ethanol in mice. We provided ad-libitum access to bottles containing either water or water and 10% ethanol to singly housed GAL5.1KO and WT male and female mice and recorded intake every 30 minutes **(Fig4 A-D).** Both male (**Fig. 4A and B**) and female (**Fig. 4C and D**) mGAL5.1KO mice consumed significantly less 10% ethanol than wild type animals when given the choice.

mGAL5.1KO mice exhibit sexual dimorphism in anxiety-like behaviour in the open field test. Previous studies have indicated a role for the galaninergic system in the pathogenesis of depression by modulating the effects of stress and anxiety [4]. To determine whether disruption of the GAL5.1 enhancer influenced anxiety levels we compared the behaviour of male and female wild type animals with that of GAL5.1KO mice using the open field test [22]. We first observed that the GAL5.1 KO reduced mobility in female animals overall but not in male mice (**Fig4 E and J**). However, male GAL5.1KO mice spent less time in the periphery (**Fig4 F and K**) whereas female mice travelled further in the periphery (**Fig 4 G and L**). The reduced length of time spent by males in the periphery was balanced against the increased time spent in the centre of the arena and an increased number of centre entries (**Fig 4 H and I**). This contrasts with female mice whose centre time and centre entry number was not affected by disruption of GAL5.1(**Fig 4 M and N**) and provides evidence of a differing role for GAL5.1 in anxiety related behaviours between males and females that parallels that observed in humans.

The EGR1 transcription factor interacts with GAL5.1 and contributes in response to PKC activation. To identify transcription factors that were involved in modulating GAL5.1 activity we undertook a bioinformatic analysis of the GAL5.1 enhancer using ENCODE data on the UCSC browser (**Fig 5A**). ENCODE identified highly conserved regions of GAL5.1 that were sensitive to DNAse1 digestion; a diagnostic of open, transcriptionally active chromatin, in several different cell lines **(Fig 5A and B).** This analysis also highlighted the presence of a highly conserved binding consensus of EGR transcription factors that lay within a DNAseI hypersensitive region (**Fig 5A and B**). Furthermore, EGR1 (AKA Zif268) is expressed in the PVN (**Fig 5C**) and up-regulated in the ARC in response to anorexia stimulated melanocortin

signalling [26]. We produced luciferase reporter constructs containing the GAL5.1(GG) enhancer and a derivative of GAL5.1(GG)lacking the conserved EGR binding site. These were transfected into a neuroblastoma cell line (SH-SY5Y) in the absence/presence of an expression vector expressing the EGR1 protein (pcDNA3-EGR1, Addgene). These experiments demonstrate that expression the EGR1 transcription factor in SH-SY5Y significantly increased GAL5.1 activity (**Fig 5D**). Critically, deletion of the conserved EGR1 binding site shown in **Fig 5B** significantly reduces the activity of GAL5.1 in the presence of EGR1 expression. We also show that treatments with 100nM of the PKC agonist phorbol ester PMA further up-regulated GAL5.1 activity in the presence of EGR1 [27-29] which was abolished in the absence of the EGR1 binding site (**Fig 5E**). Further analysis comparing the effects of EGR1 expression and PKA stimulation on the GAL5.1(GG) and GAL5.1(CA) haplotype reporter constructs showed that whilst the GG haplotype responded strongly to EGR1 expression in SH-SY5Y cells the CA haplotype did not. In addition, the PKA response of the CA haplotype was blunted in comparison to the GG haplotype of GAL5.1 (**Fig 5F**).

Discussion.

Understanding the processes that modulate preference for ethanol is particularly pressing in men where 7.6% of all male deaths globally is attributable to alcohol (4% in women) through cardiovascular disease, liver damage and cancer [30] . Anxiety remains one of the most frequent co-morbidities associated with excessive alcohol intake[31] and identifying and understanding the genomic mechanisms that link anxiety with excessive alcohol intake will be an important component in understanding and combating alcohol abuse[5]. A previous haplotype analysis of polymorphisms around the GAL gene in two different populations (Finns and Plains Native Americans) identified robust associations with alcohol intake and specific haplotypes [32]. Intriguingly, this study also uncovered evidence of a sex specific role for anxiety in modulating alcohol intake [32]. Unfortunately, the authors were unable to identify a molecular mechanism that could account for their findings. The current study took an entirely different approach and explored the problem from the initial standpoint of functionality. This was achieved by analysing the effects of polymorphisms known to change the activity of the highly conserved GAL5.1 enhancer sequence found 42kb from the GAL gene [8].

We began our analysis of a possible role for the GAL5.1 enhancer in alcohol intake by interrogating the UK Biobank cohort to determine the validity of previous conflicting smaller scale association studies [9, 10]. Our analysis supported observations by Nikolova et al (2013) who reported an association between the G-allele of GAL5.1 and increased alcohol intake in women [9] but not in men. Because anxiety has been identified as an important variable influencing alcohol intake in humans[1] and had also been identified as being an important variable in modulating the role of the GAL gene in alcohol intake[32], we carried out a deeper analysis of the rs2513280 polymorphism UK biobank by stratifying alcohol intake with anxiety. We were surprised to find a significant association between reported anxiety and drinking behaviour in men which was not observed in women. This observation was entirely consistent with previous reports of sexual dimorphism on the influence of the GAL locus on mood and alcohol intake[32].

To further explore the role of GAL5.1 in ethanol intake and anxiety we disrupted GAL5.1 in mice using CRISPR/CAS9 genome editing using a novel method of generating gRNA which greatly increased the speed and efficiency of the process. Using QrtPCR of different brain regions of GAL5.1 knock out mice we found that disruption of GAL5.1 virtually blocked expression of Gal mRNA expression in hypothalamus, amygdala, hippocampus and cortex consistent with a requirement for GAL5.1 for expression of *Gal* in these brain regions. Moreover, analysis of gene expression of four other flanking genes suggested that GAL5.1 is specific in its modulation of the expression of the *Gal* gene. However, the most striking outcome of the current study was the observation that CRISPR/CAS9 disruption of GAL5.1 significantly reduced preference for ethanol in mice demonstrating a key role for GAL5.1 in ethanol intake. Analysis of our GAL5.1 KO mice also found sexual dimorphism in the effects of GAL5.1 disruption on anxiety such that male mice demonstrated a significant reduction in anxiety-related behaviour which was not reflected in female GAL5.1KO mice. These observations are entirely consistent with our stratified analysis of the UK biobank cohort where increased anxiety was associated with increased alcohol intake in men but not in women and in previous analysis of polymorphisms around the GAL gene [32]. That patterns of male orientated anxiety and alcohol intake around a GAL5.1 haplotype should so closely parallel the ethanol intake and anxiety phenotypes identified in our analysis of the GAL5.1 disruption in mice suggests that a common, and highly conserved, mechanism governing mood and ethanol intake exists at the GAL5.1 locus. From a mechanistic standpoint it is worth considering that the observed sexual differences observed in anxiety and alcohol intake in both humans and mice may be influenced by previously observed modulation of the GAL gene promoter by the estrogen receptor [33, 34].

Several studies have identified genetic associations between polymorphisms around the human GAL locus and mood disorders [34-36]. In the current study we demonstrate that down regulation of galanin expression throughout the brain by disruption of the GAL5.1 enhancer causes significant changes in the anxiety-like behaviour of male mice subjected to the open field test. Although the current study is currently unable to confirm which areas of the brain are

responsible for this sex specific change in behaviour we show that *Gal* expression was reduced in the amygdala following disruption of GAL5.1. However, we concede that, given the known complexity of the galaninergic system in modulating mood[37], it would be premature to link the change in anxiety observed in our study with a change of expression in the amygdala. Although the specific role of *GAL* expression in the amygdala in modulating mood requires more study, the discovery of an enhancer that drives expression of *GAL* in the limbic system and which, when disrupted, alters mood affords unique possibilities for understanding the molecular mechanisms contributing to anxiety disorders that remain one of the most common co-morbidities associated with alcohol abuse[31].

Closer analysis of the GAL5.1 sequence using a combination of bioinformatics and cell co-transfection studies suggested a functional interaction between the EGR1 transcription factor and a highly conserved EGR binding site within a highly conserved region of GAL5.1. In addition, we were able to demonstrate a role for the EGR1 transcription factor in modulating previously identified PKA induced activity of GAL5.1[8]. EGR1 has a high affinity for DNA and is known to be able to bind DNA even when methylated [38]. Thus, it is entirely possible that EGR1 acts as a "pioneer factor" by being one of the first transcription factors that bind to the closed and methylated GAL5.1 locus thus activating its tissue specific activity. Interestingly, we observed that the CA allele of GAL5.1 did not respond to EGR1 expression and its PKC response was blunted. This is an interesting observation as the EGR1 binding site in GAL5.1 is 100bp from the closest of the SNPs making up the CA haplotype (rs2513281). We propose that the EGR1 protein forms part of a larger protein complex that interacts across the whole of the most conserved regions of the GAL5.1 enhancer. It is therefore possible that EGR1 binding and GAL5.1 function is dependent on binding of another, as yet unidentified, protein whose binding is interrupted by one of the allelic variants of GAL5.1 thus affecting EGR1 binding.

Conclusion. To the best of our knowledge, this is the first time that the unique combination of techniques used in our study; human association analysis, comparative genomics and CRISPR/CAS9 genome editing, have been used to establish a functional role

for a tissue-specific enhancer region in alcohol selection and mood in living animals. This study is given even greater impact by our analysis of the UK biobank cohort that demonstrates a link between increased alcohol intake and anxiety in men paralleling that seen in our CRISPR derived models. Placing the current study within a wider context of understanding the mechanistic basis of complex human disease it is clear that an important step has been made in bridging the gap between association analysis and mechanism especially in light of the fact that the majority of associated SNPs generated by GWAS are in the non-coding genome[39]. Although histone markers (H3K4me1) and GWAS can identify candidate regulatory regions affected by disease associated polymorphic variation on a genome wide level the current study serves to emphasise that there is also merit in the use of comparative genomics and functional characterisation of the cell specific activity of putative regulatory elements using CRISPR genome editing in whole animal systems.

Acknowledgements

AMcE was funded by BBSRC project grant (BB/N017544/1) and EH was funded by Medical Research Scotland (PhD-719-2013). PB and DW are funded by the Scottish Government Rural and Environment Science and Analytical Services Division to the Rowett Institute. Association studies were conducted using the UK Biobank Resource: application number 4844 and was supported by a Wellcome Trust Strategic Award 'Stratifying Resilience and Depression Longitudinally' (STRADL) (Reference 104036/Z/14/Z).

Declaration of interest. None of the authors declare any conflicts of interest

References.

1. Smith JP, Randall CL. Anxiety and alcohol use disorders: comorbidity and treatment considerations. *Alcohol Res* 2012; **34**(4)**:** 414-431.

2. Rada P, Avena NM, Leibowitz SF, Hoebel BG. Ethanol intake is increased by injection of galanin in the paraventricular nucleus and reduced by a galanin antagonist. *Alcohol* 2004; **33**(2)**:** 91-97.

3. Barson JR, Morganstern I, Leibowitz SF. Galanin and consummatory behavior: special relationship with dietary fat, alcohol and circulating lipids. *EXS* 2011; **102:** 87-111.

4. Juhasz G, Hullam G, Eszlari N, Gonda X, Antal P, Anderson IM *et al.* Brain galanin system genes interact with life stresses in depression-related phenotypes. *Proc Natl Acad Sci U S A* 2014; **111**(16): E1666-1673.

5. Turner S, Mota N, Bolton J, Sareen J. Self-medication with alcohol or drugs for mood and anxiety disorders: A narrative review of the epidemiological literature. *Depress Anxiety* 2018; **35**(9): 851-860.

6. Visel A, Bristow J, Pennacchio LA. Enhancer identification through comparative genomics. *Semin Cell Dev Biol* 2007; **18**(1): 140-152.

7. Dogan N, Wu W, Morrissey CS, Chen KB, Stonestrom A, Long M *et al.* Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin* 2015; **8:** 16.

8. Davidson S, Lear M, Shanley L, Hing B, Baizan-Edge A, Herwig A *et al.* Differential activity by polymorphic variants of a remote enhancer that supports galanin expression in the hypothalamus and amygdala: implications for obesity, depression and alcoholism. *Neuropsychopharmacology* 2011; **36**(11): 2211-2221.

9. Nikolova YS, Singhi EK, Drabant EM, Hariri AR. Reward-related ventral striatum reactivity mediates gender-specific effects of a galanin remote enhancer haplotype on problem drinking. *Genes Brain Behav* 2013; **12**(5): 516-524.

10. Richardson TG, Minica C, Heron J, Tavare J, MacKenzie A, Day I *et al.* Evaluating the role of a galanin enhancer genotype on a range of metabolic, depressive and addictive phenotypes. *Am J Med Genet B Neuropsychiatr Genet* 2014; **165B**(8): 654-664.

11. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 2017; **169**(7): 1177-1186.

12. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018; **50**(9): 1219-1224.

13. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, andMe Research Team tSUDWGotPGC, Adams MJ *et al.* Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *Am J Psychiatry* 2019; **176**(2): 107-118.

14. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562**(7726): 203-209.

15. Davis KAS, Coleman JRI, Adams M, Allen N, Breen G, Cullen B *et al.* Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. *BJPsych Open* 2018; **4**(3)**:** 83-90.

16. Harms DW, Quadros RM, Seruggia D, Ohtsuka M, Takahashi G, Montoliu L *et al.* Mouse Genome Editing Using the CRISPR/Cas System. *Curr Protoc Hum Genet* 2014; **83:** 15 17 11-27.

17. Nagy AG, M. Vinterstein, K and Behringer, R. *Manipulating the Mouse Embryo*. third edition edn. Cold Spring Harbor laboratory Press: Cold Spring Harbor, 2003.

18. Mercer JG, Lawrence CB, Atkinson T. Regulation of galanin gene expression in the hypothalamic paraventricular nucleus of the obese Zucker rat by manipulation of dietary macronutrients. *Brain Res Mol Brain Res* 1996; **43**(1-2)**:** 202-208.

19. Shanley L, Lear M, Davidson S, Ross R, Mackenzie A. Evidence for regulatory diversity and auto-regulation at the TAC1 locus in sensory neurones. *J Neuroinflammation* 2011; **8:** 10.

20. Shanley L, Davidson S, Lear M, Thotakura AK, McEwan IJ, Ross RA *et al.* Long-Range Regulatory Synergy Is Required to Allow Control of the TAC1 Locus by MEK/ERK Signalling in Sensory Neurones. *Neurosignals* 2010; **18**(3)**:** 173-185.

21. Hungund BL, Szakall I, Adam A, Basavarajappa BS, Vadasz C. Cannabinoid CB1 receptor knockout mice exhibit markedly reduced voluntary alcohol consumption and lack alcohol-induced dopamine release in the nucleus accumbens. *J Neurochem* 2003; **84**(4)**:** 698-704.

22. Crawley JN. Exploratory behavior models of anxiety in mice. *Neurosci Biobehav Rev* 1985; **9**(1)**:** 37-44.

23. Hay EA, Khalaf AR, Marini P, Brown A, Heath K, Sheppard D *et al.* An analysis of possible off target effects following CAS9/CRISPR targeted deletions of neuropeptide gene enhancers from the mouse genome. *Neuropeptides* 2016.

24. Mackenzie A, Miller KA, Collinson JM. Is there a functional link between gene interdigitation and multi-species conservation of synteny blocks? *Bioessays* 2004; **26**(11)**:** 1217-1224.

25. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature* 2012; **489**(7414)**:** 109-113.

26. Huang J, Jiang W, Yuan D. Functional profiling of immediate early gene Egr1 in an anorexic mouse model. *Mol Med Rep* 2013; **8**(5)**:** 1563-1569.

27. Halvorson LM, Kaiser UB, Chin WW. The protein kinase C system acts through the early growth response protein 1 to increase LHbeta gene expression in synergy with steroidogenic factor-1. *Mol Endocrinol* 1999; **13**(1)**:** 106-116.

28. Sun S, Ning X, Zhai Y, Du R, Lu Y, He L *et al.* Egr-1 mediates chronic hypoxia-induced renal interstitial fibrosis via the PKC/ERK pathway. *Am J Nephrol* 2014; **39**(5)**:** 436-448.

29. Yan SF, Harja E, Andrassy M, Fujita T, Schmidt AM. Protein kinase C beta/early growth response-1 pathway: a key player in ischemia, atherosclerosis, and restenosis. *J Am Coll Cardiol* 2006; **48**(9 Suppl 1)**:** A47-55.

30. Organisation WH. Global status report on alcohol and health 20142014.

31. Pacek LR, Storr CL, Mojtabai R, Green KM, La Flair LN, Alvanzo AA *et al.* Comorbid Alcohol Dependence and Anxiety Disorders: A National Survey. *J Dual Diagn* 2013; **9**(4).

32. Belfer I, Hipp H, McKnight C, Evans C, Buzas B, Bollettino A *et al.* Association of galanin haplotypes with alcoholism and anxiety in two ethnically distinct populations. *Mol Psychiatry* 2006; **11**(3)**:** 301-311.

33. Howard G, Peng L, Hyde JF. An estrogen receptor binding site within the human galanin gene. *Endocrinology* 1997; **138**(11)**:** 4649-4656.

34. Unschuld PG, Ising M, Roeske D, Erhardt A, Specht M, Kloiber S *et al.* Gender-specific association of galanin polymorphisms with HPA-axis dysregulation, symptom severity, and antidepressant treatment response. *Neuropsychopharmacology* 2010; **35**(7)**:** 1583-1592.

35. Unschuld PG, Ising M, Erhardt A, Lucae S, Kohli M, Kloiber S *et al.* Polymorphisms in the galanin gene are associated with symptom-severity in female patients suffering from panic disorder. *J Affect Disord* 2008; **105**(1-3)**:** 177-184.

36. Wray NR, Pergadia ML, Blackwood DH, Penninx BW, Gordon SD, Nyholt DR *et al.* Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry* 2012.

37. Karlsson RM, Holmes A. Galanin as a modulator of anxiety and depression and a therapeutic target for affective disease. *Amino Acids* 2006; **31**(3)**:** 231-239.

38. Hashimoto H, Olanrewaju YO, Zheng Y, Wilson GG, Zhang X, Cheng X. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev* 2014; **28**(20)**:** 2304-2313.

39. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Borglum AD, Breen G *et al.* Psychiatric Genomics: An Update and an Agenda. *Am J Psychiatry* 2018; **175**(1)**:** 15-27.

Figure Legends.

Figure 1A. Table showing interactions between different variables after analysis of the UK Biobank demonstrating a stronger than expected interaction (p=0.0008) between anxiety, sex, allelic variation at the rs2513280 locus and alcohol intake. **B.** A graph showing the linear regression line for the 3-way interaction between sex, anxiety and rs2153280, with allele count representing the number of C alleles an individual carry and the y-axis showing the log transformed AUDIT-T score.

Figure 2A and C, Dark field images of 10µm sections through **(A)** the hypothalamic region and **(C)** the amygdala region of wildtype mouse probed with S^{35} labelled *Gal* antisense RNA. **(B** and **D)** vibratome sections through **(B)** the hypothalamic and **(D)** the amygdala region of a mouse transgenic for the human GAL5.1-LacZ reporter construct (hGAL5.1-LacZ) demonstrating LacZ expression in cells of the periventricular (PVN) Dorsomedial (DMH) and arcuate (ARC) nuclei (V3, 3rd ventricle) of the hypothalamus **(A and B)** and the (MeA) medial amygdaloid nucleus of the amygdala **(C and D). E.** Comparative genomic analysis of sequence flanking the human GAL locus demonstrating the position (numbered black scale bar at top), distance (blue arrows) and depth of conservation (green peaks) of the human GAL5.1 locus (red box) relative to the GAL gene (blue box). **F and G.** Sequence data derived from PCR products of ear clip DNA (thick black bar labelled GAL5.1KO58) blasted against the mouse genome (UCSC browser). The accuracy of the disruption relative to the PAM sequence (green box) of each sgRNA demonstrates the accuracy of **(F)** 5sgRNA and **G)** 3sgRNA relative to the missing regions of the genome (thin lines and chevrons).

Figure 3A. GAL5.1 specifically regulates expression of the *Gal* gene in the hypothalamus and amygdala but not flanking genes . A. Bar graph showing relative levels of Gal expression as measured by QrtPCR in different brain areas (Hypo, hypothalamus; Amyg, amygdala; Hipp, hippocampus, Cort; cortex) in wild type (WT) and mGAL5.1KO animals. **B.** Scatterplots demonstrating QrtPCR analysis of Gal mRNA expression in RNA derived from hypothalamus, amygdala, hippocampus and cortex on total RNA derived from

wild type (WT) or GAL5.1 knockout (mGAL5.1KO) animals(**;p<0.01). **C.** Scale diagram (UCSC genome browser) representing genes surrounding the mGAL5.1 enhancer (red box). Exons are displayed as thick blue lines and introns by thin blue lines punctuated by chevrons denoting the direction of transcription. Genomic coordinates on mouse chromosome 19 are highlighted by black perpendicular lines. **D.** Scatterplots showing QrtPCR analysis of cDNA derived from total hypothalamic RNA comparing mRNA expression of *Lrp5, Ppp6r3, Gal, Mtl5* and *Cpt1A* in wild type and GAL5.1KO animals normalised against the expression of the *Nono* gene (y-axis). (****, p<0.001, n.s., no significance).

Figure 4. GAL5.1 regulates ethanol intake and modulates anxiety like behaviour in a sex specific manner. Graphs demonstrating **(A and C)** fluid intake recorded every 30 mins of singly housed **(A)** male and **(C)** female mice offered the choice of water or 10% ethanol showing intake of the wild type (black data points) or mGAL5.1KO animals (red data points) over 10 days. **B and D** scatter plot graphs showing percentage fluid intake over 10 days in **(C)** male and **(D)** female singly housed wild type and mGAL5.1KO mice (error bars= SEM; n=7, n.s. not significant; p<0.05, **; p<0.01, ***; p<0.005. ****, p<0.0001). **E-N** scatter plots showing results of open field test performed in **(E-I)** male and **(J-N)** female mice where **E** and **J** shows the distance each mouse travelled over 300 seconds, **F** and **K** shows the time each mouse spent in the peripheral area of the test field, **G** and **L** shows the distance travelled within the peripheral area, **H** and **M** shows the number of times each mouse entered the centre area and **I** and **N** shows the total length of time each mouse stayed in the central area.(error bars= SEM; n=12, n.s. not significant, *; p<0.05, **; p<0.01).

Figure 5. EGR1 modulates the PKA response of GAL5.1. **A.** UCSC browser output of 800bp surrounding the GAL5.1 enhancer showing regions of DNAseI hypersensitivity (**Ai**; filled grey bars), conserved transcription factor binding consensus sequences (**Aii**; filled black boxes) and degree (**Aiii;** blue peaks) and depth (A**iii;** green lines) of sequence conservation. **B.** UCSC output showing hypersensitivity sites (grey bars), conserved EGR1 consensus sequences (black bars) and degrees of conservation (Blue peaks). Multiz alignment of

conserved consensus sequences demonstrates degrees of conservation at the base pair level where dots represent identical base pair to human. **C.** Allen brain atlas output showing expression of EGR1 mRNA within the periventricular nucleus (PVN) of the hypothalamus. **D.** A bar graph demonstrating dual luciferase data derived from SH-SY5Y cells transfected with different combinations of pcDNA3 (empty expression vector), pLuc (empty reporter vector), pcDNA-EGR1 (Expression vector expressing EGR1 transcription factor), pLuc-GAL5.1 (reporter vector containing the GAL5.1 enhancer), pLuc-Gal∆EGR (reporter construct containing the GAL5.1 enhancer lacking the EGR binding consensus shown in **B**). **E.** A bar graph showing dual luciferase data of SH-SY5Y cells transfected with the expression and reporter constructs described above and cultured in the presence or absence of 100nM PMA. All cultures were treated with the same concentration of the vehicle DMSO. (n=4, ns; no significance, ********:p<0.001). **F.** A bar graph comparing the effects of co-transfection of an EGR1 expressing plasmid (pcDNA-EGR1) and/or PMA treatment on cells transfected with a renilla luciferase expressing plasmid and a firefly luciferase reporter constructs containing either the GG or CA haplotypes of the GAL5.1 enhancer (n=4, ns; no significance, ********:p<0.001).

Figure 1

A

Variable	Estimate	S.E.	p-value
Sex (male)	0.34	0.005	**$< 2 \times 10^{-16}$**
rs2513280	0.003	0.006	0.58
Anxiety	0.03	0.006	**3.7×10^{-6}**
Sex (male) * anxiety	-0.005	0.01	0.60
Sex (male) * rs2513280	-0.01	0.009	0.12
Anxiety * rs2513280	-0.02	0.01	0.10
Sex (male) * anxiety * rs2513280	0.06	0.02	**0.0008**

B

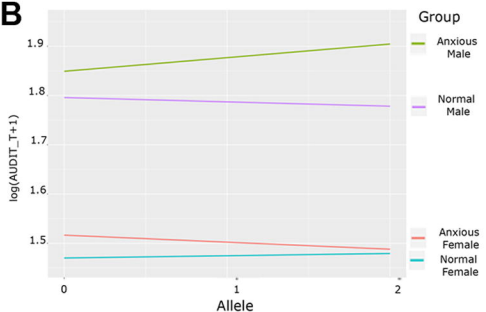


Figure 2

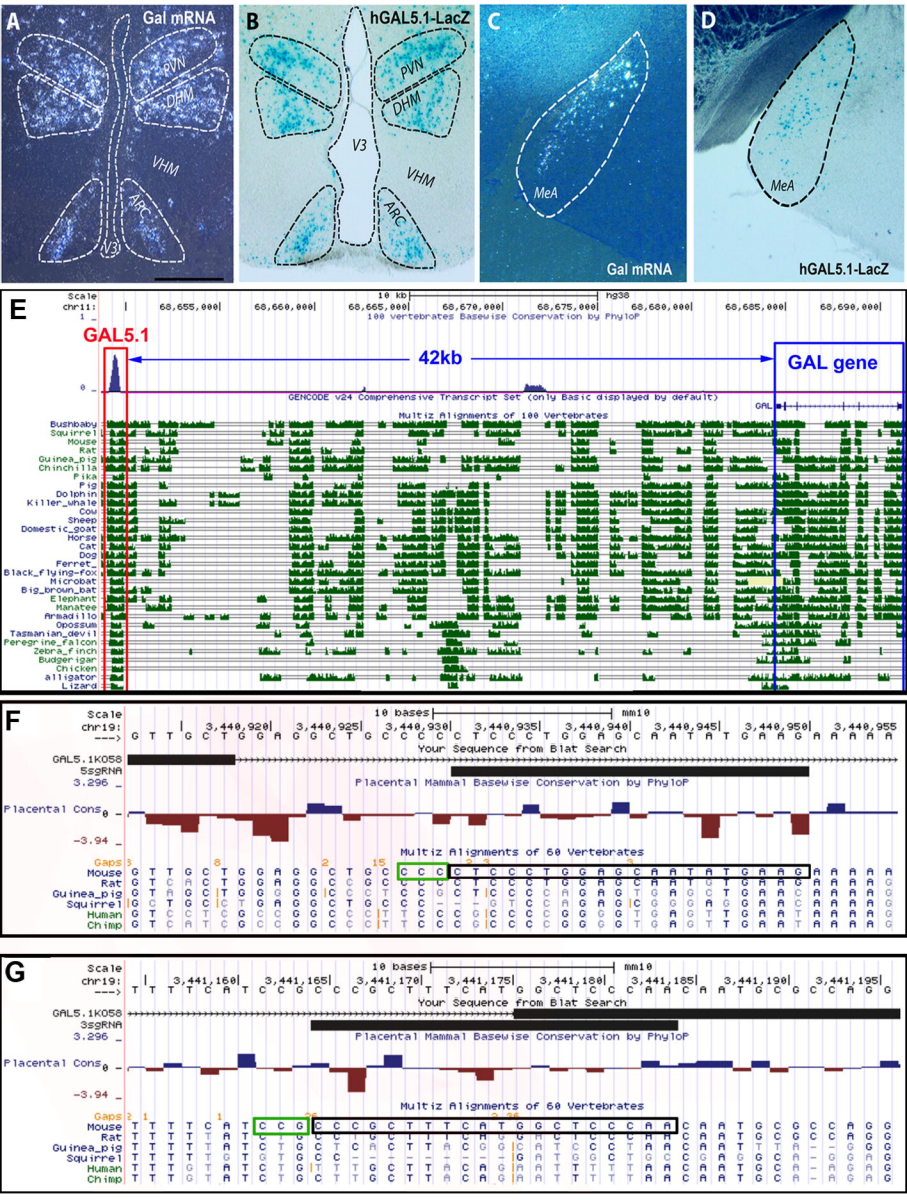

A
Gal mRNA
PVN
DHM
VHM
ARC
V3
B
hGAL5.1-LacZ
PVN
DHM
V3
VHM
ARC
C
MeA
Gal mRNA
D
MeA
hGAL5.1-LacZ
E
GAL5.1
42kb
GAL gene
GENCODE V24 Comprehensive Transcript Set (only Basic displayed by default)
Multiz Alignments of 100 Vertebrates
F
Your Sequence from Blat Search
GAL5.1K058
5sgRNA
Placental Mammal Basewise Conservation by PhyloP
Multiz Alignments of 60 Vertebrates
Mouse
Rat
Guinea_pig
Squirrel
Human
Chimp
G
Your Sequence from Blat Search
GAL5.1K058
3sgRNA
Placental Mammal Basewise Conservation by PhyloP
Multiz Alignments of 60 Vertebrates
Mouse
Rat
Guinea_pig
Squirrel
Human
Chimp

Figure 3

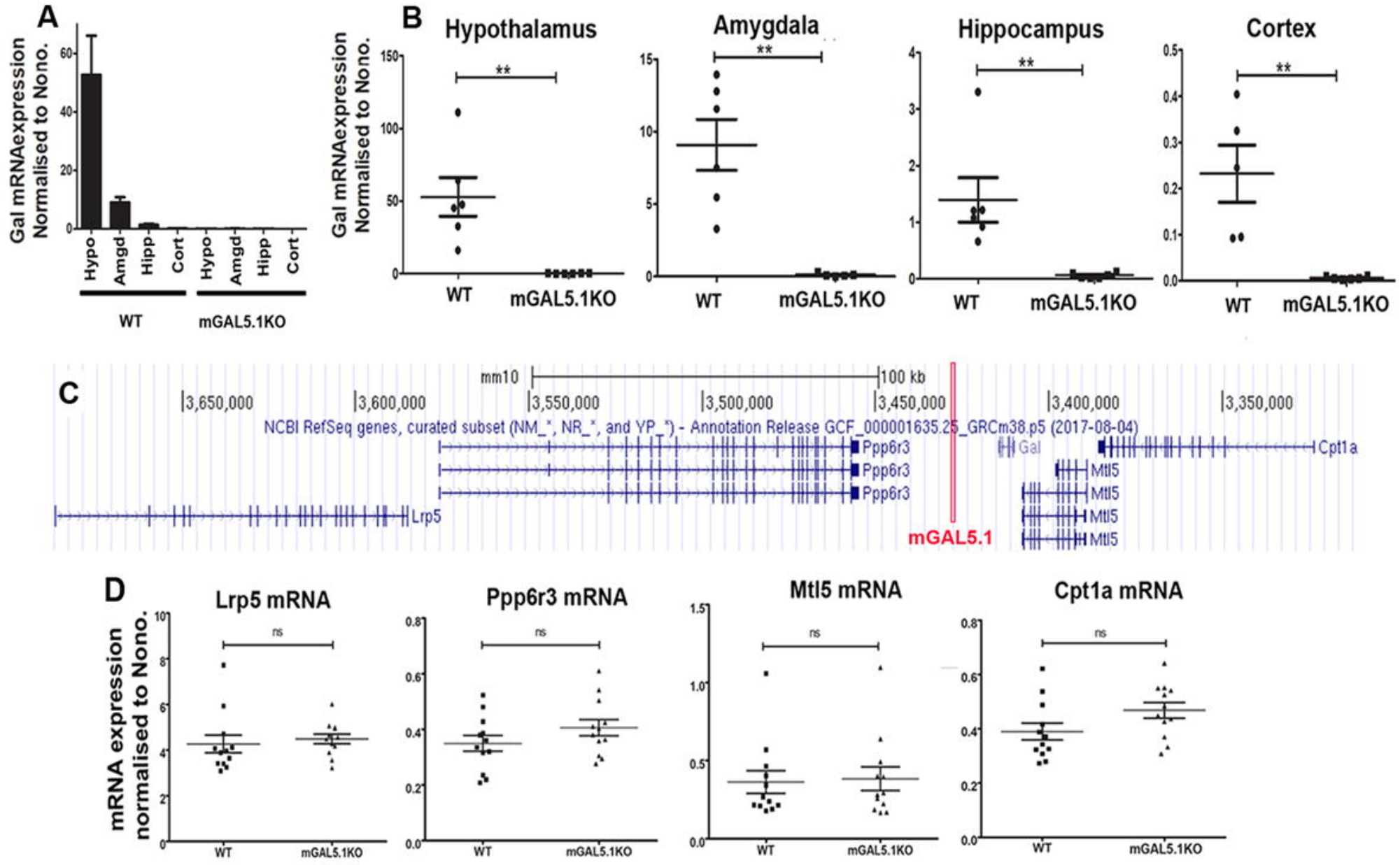

A
Gal mRNAexpression Normalised to Nono.
Hypo
Amgd
Hipp
Cort
Hypo
Amgd
Hipp
Cort
WT
mGAL5.1KO
B
Gal mRNAexpression Normalised to Nono.
Hypothalamus
Amygdala
Hippocampus
Cortex
**
WT
mGAL5.1KO
C
mm10
100 kb
3,650,000
3,600,000
3,550,000
3,500,000
3,450,000
3,400,000
3,350,000
NCBI RefSeq genes, curated subset (NM_*, NR_*, and YP_*) - Annotation Release GCF_000001635.25_GRCm38.p5 (2017-08-04)
Ppp6r3
Ppp6r3
Ppp6r3
Lrp5
Gal
Cpt1a
Mtl5
Mtl5
Mtl5
Mtl5
mGAL5.1
D
mRNA expression normalised to Nono.
Lrp5 mRNA
Ppp6r3 mRNA
Mtl5 mRNA
Cpt1a mRNA
ns
WT
mGAL5.1KO

Figure 4

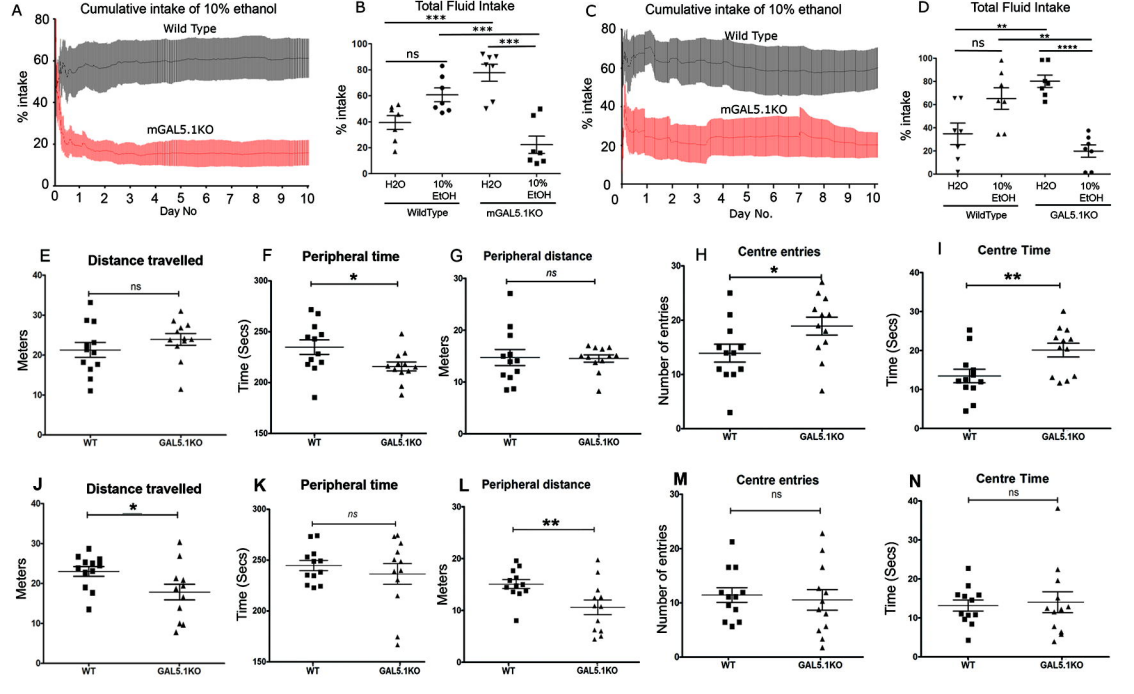
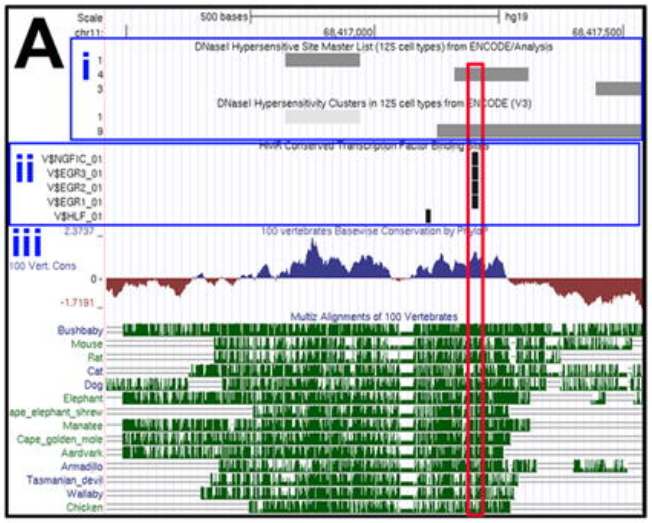
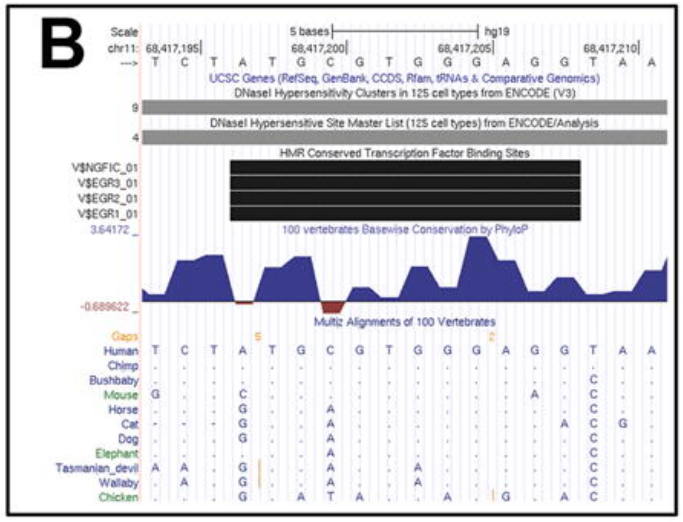
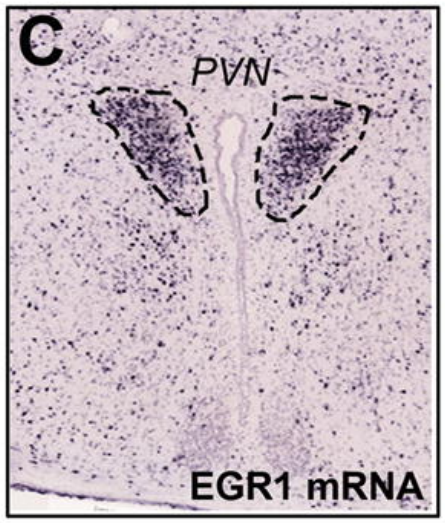
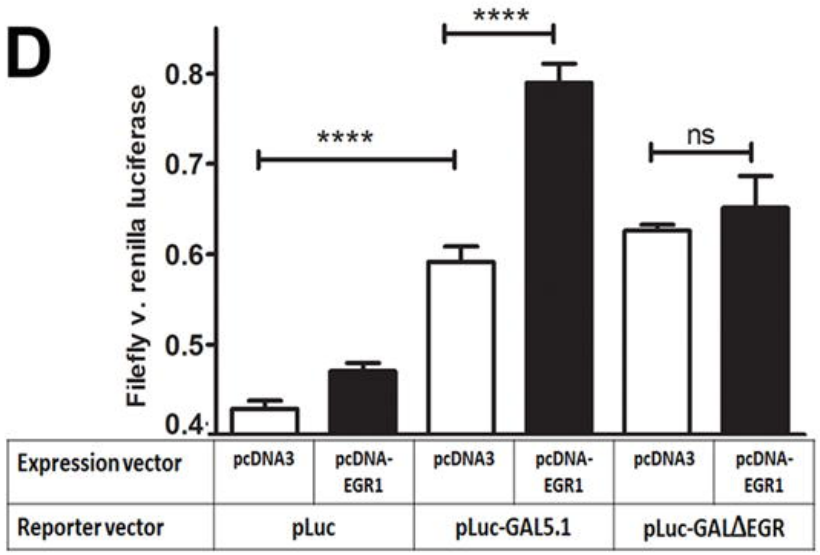

A
Cumulative intake of 10% ethanol
Wild Type
mGAL5.1KO
% intake
Day No
B
Total Fluid Intake
ns
***
***
***
% intake
H2O
10% EtOH
H2O
10% EtOH
WildType
mGAL5.1KO
C
Cumulative intake of 10% ethanol
Wild Type
mGAL5.1KO
% intake
Day No.
D
Total Fluid Intake
ns
**
**
****
% intake
H2O
10% EtOH
H2O
10% EtOH
WildType
GAL5.1KO
E
Distance travelled
ns
Meters
WT
GAL5.1KO
F
Peripheral time
*
Time (Secs)
WT
GAL5.1KO
G
Peripheral distance
ns
Meters
WT
GAL5.1KO
H
Centre entries
*
Number of entries
WT
GAL5.1KO
I
Centre Time
**
Time (Secs)
WT
GAL5.1KO
J
Distance travelled
*
Meters
WT
GAL5.1KO
K
Peripheral time
ns
Time (Secs)
WT
GAL5.1KO
L
Peripheral distance
**
Meters
WT
GAL5.1KO
M
Centre entries
ns
Number of entries
WT
GAL5.1KO
N
Centre Time
ns
Time (Secs)
WT
GAL5.1KO

Figure 5

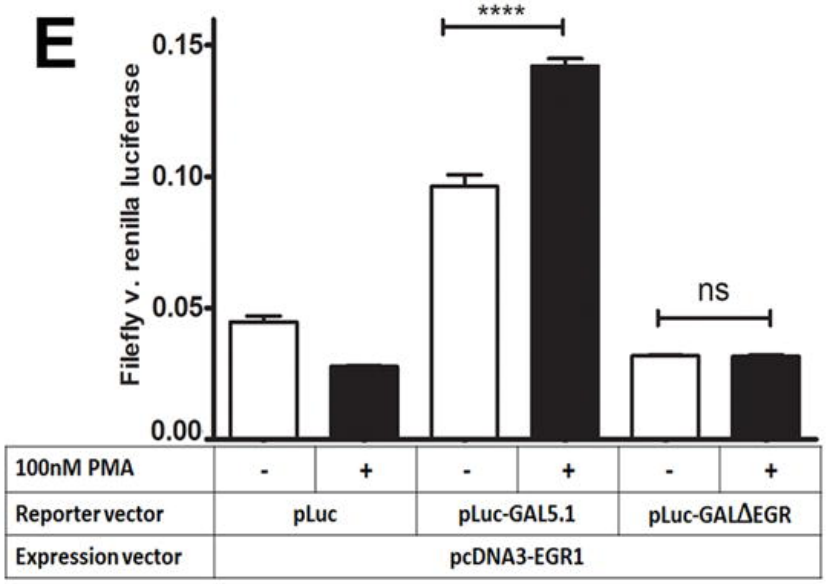




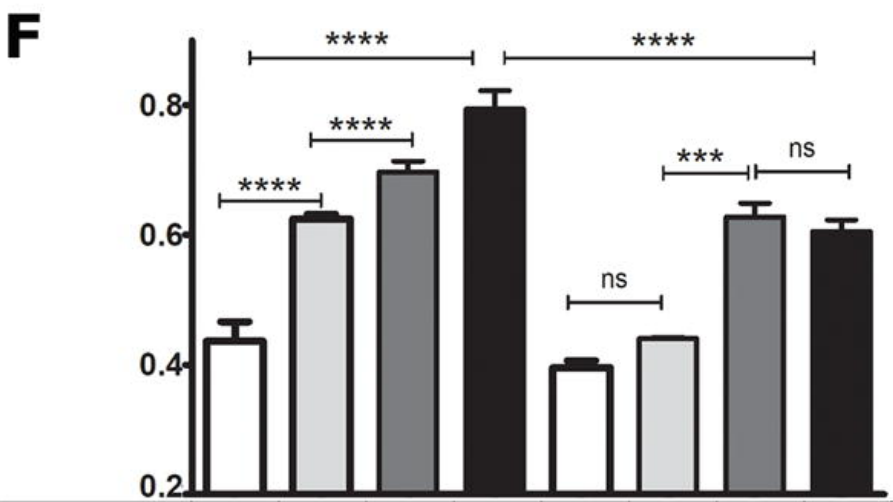




Expression vector	pcDNA3	pcDNA-EGR1	pcDNA3	pcDNA-EGR1	pcDNA3	pcDNA-EGR1
Reporter vector	pLuc		pLuc-GAL5.1		pLuc-GALΔEGR	



100nM PMA	-	+	-	+	-	+
Reporter vector	pLuc		pLuc-GAL5.1		pLuc-GALΔEGR	
Expression vector	pcDNA3-EGR1					



Expression vector	pcDNA 3	pcDNA-EGR1	pcDNA 3	pcDNA-EGR1	pcDNA 3	pcDNA-EGR1	pcDNA 3	pcDNA-EGR1
100nM PMA	-		+		-		+	
Reporter vector	pLucGAL5.1(GG)				pLucGAL5.1(CA)			